Error Analysis of Low-rank Three-Way Tensor Factorization Approach to Blind Source Separation

Ivica Kopriva, Jean-Philip Royer, Nadège Thirion-Moreau, Pierre Comon

ERROR ANALYSIS OF LOW-RANK THREE-WAY TENSOR FACTORIZATION APPROACH TO BLIND SOURCE SEPARATION

Ivica Kopriva[1], Jean-Philip Royer[2], Nadège Thirion-Moreau[2], Pierre Comon[3]

[1] Division for Laser and Atomic R&D, Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia
[2] Aix-Marseille Université, CNR, ENSAM, LSIS, UMR 7296, 13397 Marseille, France and Université de Toulon, CNRS, LSIS, UMR 7296, 83957, La Garde, France
[3] GIPSA-Lab, CNRS UMR5216, Grenoble Campus, BP.46 F-38402 St Martin d'Heres Cedex, France

ABSTRACT

In tensor factorization approach to blind separation of multidimensional sources two formulas for calculating the source tensor have emerged. In practice, it is observed that these two schemes exhibit different levels of robustness against perturbations of the factors involved in the tensor model. Motivated by both practical reasons and the will to better figure this out, we present error analyses in source tensor estimation performed by low-rank factorization of three-way tensors. To that aim, computer simulations as well as the analytical calculation of the theoretical error are carried out. The conclusions drawn from these numerical and analytical error analyses are supported by the results obtained thanks to the tensor factorization based blind decomposition of an experimental multispectral image of a skin tumor.

Index Terms— Tensor models, multidimensional signal, blind source separation

1. INTRODUCTION

Great majority of data acquired in contemporary medical imaging, remote sensing and/or spectroscopy are three-way or three-dimensional (3D) tensors. Few examples include MultiSpectral/HyperSpectral Image (MSI/HSI), [1-4], multispectral Magnetic Resonance Image (mMRI), [5], functional MRI, [6], multi-phase computed tomography image, [7], and fluorescence spectroscopy, [8]. Their blind (a.k.a. unsupervised) decomposition is receiving increasing attention since profiles (temporal, spectral, density, etc.) of the objects (materials, organs, tissues, compounds) present in the data do not need to be known to be recovered. However, a great majority of blind decomposition methods rely on two-dimensional representation of the multichannel data, [1, 2, 5-7], even when these data are inherently 3D tensors. Moreover, in fluorescence spectroscopy, [8], data tensor obeys the canonic polyadic decomposition (CPD) tensor model (a.k.a. PARAllel FACtor analysis – CP/PARAFAC), [9, 10]. Tensor factorization (TF) approach to Blind Source Separation (BSS) is justified by the fact that hard constraints, such as sparseness or statistical independence, necessary in matrix factorization approaches to BSS, can be substituted by multidimensionality, [11, 12]. In this regard, it has been demonstrated in [3] that Nonnegative TF (NTF) based on α-divergence cost function, a.k.a α-NTF [13, 14], is capable to perform blind MSI decomposition by using nonnegativity constraint only. However, as briefly mentioned in [3], the TF approach to BSS yields two formulas for calculating the estimate of the source tensor relying on all or a part of the estimated factors of the 3D tensor model. Thus, the goal of this paper is to present an error analysis that will help make the right choice of formula according to a specific application scenario. To this end, both numerical and analytical error analyses are presented for two tensor models: the Tucker3 model [15] and the CPD model. The Tucker3 model is treated even though formal proof is still missing regarding conditions necessary for achieving a unique decomposition. That is because there is a vast number of papers demonstrating meaningfulness of results obtained by factorization of data tensors according to this Tucker model among which [3, 4, 14]. The considered error analysis is related to low-rank approximations of 3D tensors denoted by $\underline{\mathbf{X}} \in \mathbb{R}^{\alpha \times \beta \times \gamma}$, but computer simulations will be performed for nonnegative tensors to stick to the outlined practical examples. When $\underline{\mathbf{X}}$ represents a multichannel image, γ stands for the number of channel images whose size is then $\alpha \times \beta$ pixels. Low-rank constraint implies that the rank, r, of $\mathbf{X}_{(3)} \in \mathbb{R}^{\gamma \times \alpha\beta}$ matrix, obtained by mode-3 unfolding of $\underline{\mathbf{X}}$, is constrained by the number of channels, i.e. $r \leq \gamma$. Since r matches the number of objects present in the data tensor, such a low-rank constraint is meaningful and justified from an experimental point of view. Even though analytical error analyses in sections 3 and 4 are related to 3D tensors, it is obvious that they can be

extended to higher-dimensional tensors. However, the algebraic complexity of expressions derived for additive noise type of error analysis, presented in section 4, would significantly grow with the tensor order. Nevertheless, this type of error analysis supplements the results of the Cramer-Rao Lower Bound (CRLB) analysis presented in [16]. While in [16] error bounds on parameter estimates of the CPD model are derived upon a white Gaussian noise assumption, the additive noise type of error analysis presented in this article is distribution invariant.

2. PROBLEM FORMULATION

Given a 3D tensor $\underline{\mathbf{X}}$, the Tucker3 decomposition consists of the following triadic or trilinear decomposition:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} + \delta\underline{\mathbf{X}}$$

where $\underline{\mathbf{G}}$ denotes the core tensor, of smaller size than $\underline{\mathbf{X}}$, $\delta\underline{\mathbf{X}}$ denotes the additive noise or measurement error and $\bullet_n$ stands for the *n*-mode product of a tensor with a matrix, i.e. contraction over the *n*th tensor index and 2nd matrix index. This decomposition is not unique. If the core tensor is constrained to be diagonal ($\underline{\mathbf{G}} = \underline{\mathbf{\Lambda}}$), then the so-called CP decomposition is unique under known sufficient conditions (cf. section 4):

$$\underline{\mathbf{X}} = \underline{\mathbf{\Lambda}} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C} + \delta\underline{\mathbf{X}}$$

In the remainder, sizes of matrices **A**, **B** and **C** are respectively $\alpha \times r$, $\beta \times r$ and $\gamma \times r$. Due to the reasons outlined in section 1 we impose the low-rank constraint on *r*, i.e. $r \leq \gamma$. In this framework, we are mainly interested in the calculation of a signal tensor, which can be expressed in two different ways:

$$\underline{\mathbf{S}}^{dir} = \underline{\mathbf{G}} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \qquad (1)$$

or

$$\underline{\mathbf{S}}^{inv} = \underline{\mathbf{X}} \bullet_3 \mathbf{C}^{\dagger} \qquad (2)$$

which we shall, respectively, call the "direct" and "inverse" expressions, and where $(.)^{\dagger}$ stands for the pseudo-inverse operator. Signal tensor of size $r \times \alpha \times \beta$ contains *r* intensity images of size $\alpha \times \beta$ pixels that indicate spatial locations of *r* objects present in the image. The goal of this paper is to determine how to choose between direct and inverse formulas. To assist in making such a decision, the sections 3 and 4 are dedicated to numerical and analytical error analyses in the estimation of $\underline{\mathbf{S}}^{dir}$ and $\underline{\mathbf{S}}^{inv}$ for Tucker3 and CPD tensor models. Conclusions drawn from these analyses are supported, at the end of section 3, by the results of the blind decomposition of an experimental multispectral image with a known ground truth.

3. NOISE ON LOADING MATRICES

This analysis of robustness is based on the first order perturbation of expressions (1) and (2). On one hand, we have:

$$\delta\underline{\mathbf{S}}^{dir} \approx \delta\underline{\mathbf{G}} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} + \underline{\mathbf{G}} \bullet_1 \delta\mathbf{A} \bullet_2 \mathbf{B} + \underline{\mathbf{G}} \bullet_1 \mathbf{A} \bullet_2 \delta\mathbf{B} \qquad (3)$$

and on the other hand, by exploiting results derived in [16], we find that:

$$\delta\underline{\mathbf{S}}^{inv} \approx -\underline{\mathbf{X}} \bullet_3 \begin{bmatrix} \mathbf{C}^{\dagger}\delta\mathbf{C}\mathbf{C}^{\dagger} + \mathbf{C}^{\dagger}\mathbf{C}^{\dagger T}\delta\mathbf{C}^{T}\left(\mathbf{I}_{\gamma} - \mathbf{C}\mathbf{C}^{\dagger}\right) \\ + \left(\mathbf{I}_{r} - \mathbf{C}^{\dagger}\mathbf{C}\right)\delta\mathbf{C}^{T}\mathbf{C}^{\dagger T}\mathbf{C}^{\dagger} \end{bmatrix} \qquad (4)$$

where I_{ε} denotes the $\varepsilon \times \varepsilon$ identity matrix, and $(.)^T$ stands for the transpose operator. It reduces to:

$$\delta\underline{\mathbf{S}}^{inv} \approx -\underline{\mathbf{X}} \bullet_3 \left[\mathbf{C}^{-1}\delta\mathbf{C}\mathbf{C}^{-1}\right] \qquad (5)$$

if the matrix **C** happens to be square invertible. In this kind of error analysis, it is assumed that the loading matrices are estimated thanks to some known TF algorithm ($\delta\underline{\mathbf{X}}$ is tensor with all entries equal to zero), and the sensitivity of $\underline{\mathbf{S}}^{dir}$ and $\underline{\mathbf{S}}^{inv}$ is evaluated when $\underline{\mathbf{G}}$, **A**, **B** and **C** are perturbed by independent and identically distributed additive noises. To some extent, this could represent rounding errors due to computation or storage. It can also represent errors due to convergence to local minima. Of course, a crucial point is the conditioning of matrix **C**, which controls the amplitude of the perturbation affecting $\underline{\mathbf{S}}^{inv}$, and hence determines which of the solutions (direct or inverse) is more robust. As it can be observed in Figure 1, when matrix **C** is *sufficiently well conditioned*, the inverse formula is more accurate. In this figure, $\log_{10}$ of the ratio of the Frobenius norms of error tensors (3) and (5) is shown versus the condition number of the **C** matrix. The original tensor $\underline{\mathbf{X}}$ was of size 856×1144×3. Entries of the loading matrices **A**, **B** and **C** and core tensor $\underline{\mathbf{G}}$ were drawn independently according to nonnegative uniform distribution, where intrinsic dimensions were equal to *r*=3. The 3×3 **C** matrix was generated with controlled condition number between 2 and 20 in steps of 1. Perturbations of $\underline{\mathbf{G}}$, **A**, **B** and **C** were generated as nonnegative uniform noise with Frobenius norms equal to 0.1%, 1% and 10% of the Frobenius norms of $\underline{\mathbf{G}}$, **A**, **B** and **C**, respectively. For each value of the condition number of **C**, one hundred cases were generated. Size of the original tensor, loading matrices and core tensor as well as type of perturbation distribution comply with the typical MSI case but, if necessary, could be selected differently. Figure 1 shows the mean values as well as the minimal and maximal values of the $\log_{10}$ of the ratio. The inverse formula yields overwhelmingly smaller error than direct formula when the condition number of **C** matrix is less than or equal to 16.

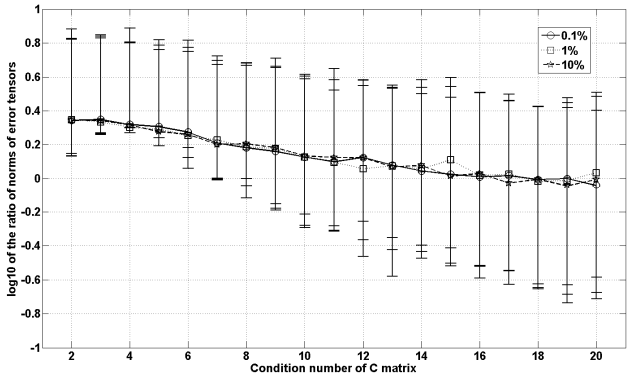

Fig. 1. Log_{10} of the ratio of the Frobenius norms of error tensors (3) and (5). Errors are due to uniform random perturbations in the loading factors of the Tucker3 tensor model in the amounts of 0.1%, 1% and 10% of the Frobenius norms of the true values of loading factors.

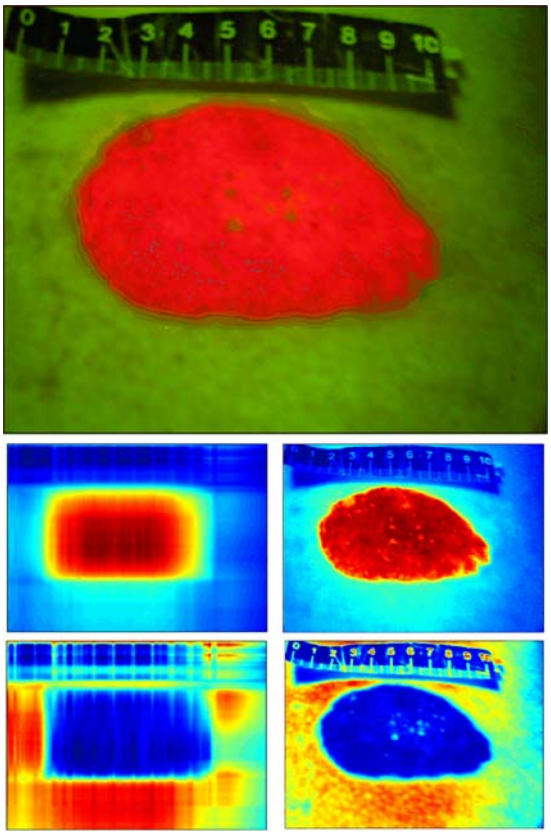

Fig. 2. (color online). Top row: experimental RGB fluorescent image of a skin tumor that stands for measurement tensor $\mathbf{X}$ of dimensions 856×1144×3. Mid row: intensity maps of tumor component. Left direct formula (1); right: inverse formula (2). Bottom row: intensity maps of background component. Left direct formula (1); right: inverse formula (2). Intensity maps are scaled to [0, 1] interval and shown in pseudo-color such that dark red indicates that component is present with probability 1, while dark blue indicates that component is present with probability 0.

To further support the conclusions drawn from this numerical and analytical error analyses, we have repeated the experiment previously reported in [3]. It is related to the unsupervised decomposition of fluorescent RGB image of a skin tumor (basal cell carcinoma) by means of the α-NTF algorithm [13] (α=0.1 was chosen in this case). Here, α-divergence is just a choice and other cost functions could be used as well. Results are shown in Figure 2. The ground truth is simple and visible on the RGB image itself which is shown in top row of Figure 2. The image contains fluorescent tumor component in red color and background component (composed of surrounding healthy skin and the ruler) in green and black colors, that is the $\mathbf{C}$ matrix is 3×2 matrix. Although result obtained by inverse formula (2) is better, it can be seen from Figure 2 that direct formula (1) also yields result that is meaningful. Thus, if the conditioning of the $\mathbf{C}$ matrix happens to be poor (it could be due to the existence of spectrally very close or even similar objects in $\underline{\mathbf{X}}$) direct formula (1) can be useful.

4. NOISE ON MEASUREMENT TENSOR

In the second analysis, only tensor $\underline{\mathbf{X}}$ is perturbed and corresponding perturbations of factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ as well as tensor $\underline{\mathbf{\Lambda}}$ are calculated. Hence, the latter perturbations are not independent nor of same variance anymore. It is clear that this analysis can be carried out only when the decomposition is unique, which discards the Tucker3 decomposition. In order to carry out the computer simulations, it is now necessary to have the expressions of $\delta\mathbf{A}$, $\delta\mathbf{B}$, $\delta\mathbf{C}$ and $\delta\underline{\mathbf{\Lambda}}$ as a function of $\delta\underline{\mathbf{X}}$. For this purpose, we consider again the first order expansion of the CPD of $\underline{\mathbf{X}}$, which yields:

$$\delta\mathbf{X}_{(3)} \approx \delta\mathbf{C}\mathbf{\Lambda}\left[\mathbf{B}\odot\mathbf{A}\right]^{\mathrm{T}} + \mathbf{C}\delta\mathbf{\Lambda}\left[\mathbf{B}\odot\mathbf{A}\right]^{\mathrm{T}} + \mathbf{C}\mathbf{\Lambda}\left[\delta\mathbf{B}\odot\mathbf{A}\right]^{\mathrm{T}} + \mathbf{C}\mathbf{\Lambda}\left[\mathbf{B}\odot\delta\mathbf{A}\right]^{\mathrm{T}}$$

where $\odot$ denotes the column-wise Kronecker product, often referred to as Khatri-Rao product, see e.g. page 30 in [18]. $\mathbf{\Lambda}$ is a $r\times r$ diagonal matrix containing the entries of the diagonal tensor $\underline{\mathbf{\Lambda}}$, if r denotes the rank of $\underline{\mathbf{X}}$. This expansion is based on the third mode unfolding of $\underline{\mathbf{X}}$ i.e. $\mathbf{X}_{(3)} = \mathbf{C}\mathbf{\Lambda}\left[\mathbf{B}\odot\mathbf{A}\right]^{\mathrm{T}}$. The derived expansion is a linear system and can be rewritten in a more convenient form, by defining the two following vectors:

$$\delta\mathbf{z} \triangleq \begin{bmatrix} vec\{\delta\mathbf{A}\} \\ vec\{\delta\mathbf{B}\} \\ vec\{\delta\mathbf{C}\} \\ vecd\{\delta\mathbf{\Lambda}\} \end{bmatrix} \quad \text{and} \quad \delta\mathbf{x} \triangleq vec\left\{\delta\mathbf{X}_{(3)}\right\}$$

where $vec\{.\}$ and $vecd\{.\}$ respectively mean:

$$vec\{\delta\mathbf{A}\} \triangleq \begin{bmatrix} \delta\mathbf{a}_1 \\ ... \\ \delta\mathbf{a}_r \end{bmatrix}, \quad vec\{\delta\mathbf{B}\} \triangleq \begin{bmatrix} \delta\mathbf{b}_1 \\ ... \\ \delta\mathbf{b}_r \end{bmatrix}, \quad vec\{\delta\mathbf{C}\} \triangleq \begin{bmatrix} \delta\mathbf{c}_1 \\ ... \\ \delta\mathbf{c}_r \end{bmatrix}$$

and $vecd\{\delta\mathbf{\Lambda}\} \triangleq \begin{bmatrix} \delta\lambda_{11} \\ ... \\ \delta\lambda_{rr} \end{bmatrix}$ and $\mathbf{a}_1,\ldots,\mathbf{a}_r$ stand for the r columns of the loading matrix $\mathbf{A}$. Then, we have to solve for $\delta\mathbf{z}$ the linear system $\mathbf{M}\delta\mathbf{z}=\delta\mathbf{x}$, where the 4-block matrix $\mathbf{M}$ is written as ("|" standing for the separation between the different blocks):

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{K}_{\alpha\beta,r}\left(\mathbf{C\Lambda}\otimes\mathbf{I}_{\alpha\beta}\right)D(\mathbf{B}) & \mathbf{K}_{\alpha\beta,r}\left(\mathbf{C\Lambda}\otimes\mathbf{I}_{\alpha\beta}\right)D(\mathbf{A}) \\ (\mathbf{B}\odot\mathbf{A})\mathbf{\Lambda}\otimes\mathbf{I}_\gamma & (\mathbf{B}\odot\mathbf{A})\odot\mathbf{C} \end{bmatrix} \quad (6)$$

where $D(\mathbf{A}) = Diag\{\mathbf{I}_\beta\otimes\mathbf{a}_1,...,\mathbf{I}_\beta\otimes\mathbf{a}_r\}$ and $D(\mathbf{B}) = Diag\{\mathbf{b}_1\otimes\mathbf{I}_\alpha,...,\mathbf{b}_r\otimes\mathbf{I}_\alpha\}$ are block-diagonal matrices of size $\alpha\beta r\times\beta r$ and $\alpha\beta r\times\alpha r$, respectively, and $\otimes$ stands for the Kronecker product. Note that there probably exist other more compact ways of writing $D(\mathbf{A})$ and $D(\mathbf{B})$, for instance by using the selection matrix $\mathbf{E}$ satisfying $(\mathbf{B}\odot\mathbf{A})=(\mathbf{B}\otimes\mathbf{A})\mathbf{E}$, [19], but this would be computationally very suboptimal. This matrix $\mathbf{M}$ is of size $\alpha\beta\gamma\times(\alpha+\beta+\gamma+1)r$. Due to low-rank constraint, $r\le\gamma$, and due to the fact that γ stands for number of channels which is low (γ=3 in case of RGB image analyzed in section 3) matrix $\mathbf{M}$ has less columns than rows; in other words, we have more equations than unknowns. Eventually, every matrix perturbation can be obtained from $\delta\underline{\mathbf{X}}$. Figure 3 shows $\log_{10}$ of the ratio of the Frobenius norms of error tensors (3) and (5) versus the condition number of the $\mathbf{C}$ matrix. The original tensor $\underline{\mathbf{X}}$ was of the size 50×50×3. For each realization of $\underline{\mathbf{X}}$ entries of loading matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ were drawn from nonnegative uniform distributions, where dimensions were equal to r=3. The core tensor $\underline{\mathbf{\Lambda}}$ was generated with nonnegative uniformly distributed values on diagonal. Please note that, if necessary, tensors with positive and negative entries could be generated as well. The 3×3 $\mathbf{C}$ matrix was generated with controlled condition number between 2 and 20 in steps of 1. Entries of tensor perturbation $\delta\underline{\mathbf{X}}$ were drawn independently according to a nonnegative uniform distribution, whereas Frobenius norm of $\delta\underline{\mathbf{X}}$ was determined from a predefined signal-to-noise ratio (SNR) in dB as: $SNR = 20\log_{10}\left(\|\underline{\mathbf{X}}\|/\|\delta\underline{\mathbf{X}}\|\right)$. The numerical evaluation scenario is consistent with the one used in section 3 to obtain Figure 1. Figure 3 shows the mean values as well as the minimal and maximal values of the $\log_{10}$ of the ratio. The inverse formula yields overwhelmingly smaller error than direct formula when condition number of $\mathbf{C}$ matrix is less than or equal to 8. The fact that inverse formula is more sensitive to measurement noise than noise in loading factors is expected since it increases noise via $\mathbf{C}^\dagger$. Results presented in this section supplement the one related to Cramer-Rao lower bound (CRLB) analysis presented in [16]. While CRLB predicts error bounds on parameter of the CPD model under a white Gaussian noise assumption, the error analysis presented herein can be performed for arbitrary distribution of the additive noise.

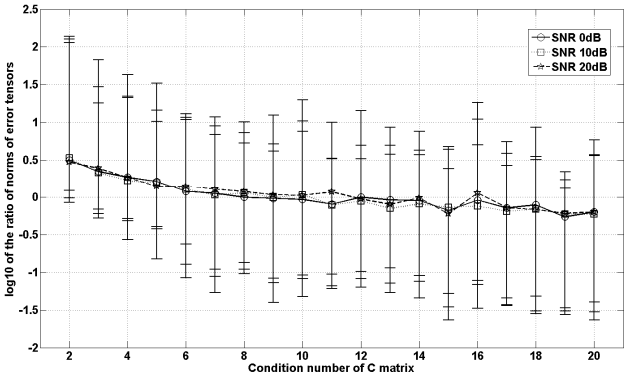

Fig. 3. Log_{10} of the ratio of the Frobenius norms of error tensors (3) and (5). Errors are due to uniform random perturbations in the measurement tensor based on the CPD tensor model.

5. CONCLUSION

In factorization of three-way tensors direct and inverse formulas for calculating source tensor emerge. In case of measurement noise and mode-*3* dimension equal to 3, inverse formula is better when condition number of mode-*3* loading matrix is smaller than or equal to 8. If errors are due to perturbations in loading matrices inverse formula is better when condition number of the mode-*3* loading matrix is smaller than or equal to 16.

APPENDIX

Eq. (6) was obtained thanks to a certain number of identities. Some of them are recalled below:

$$vec\{\mathbf{UVW}\} = \left(\mathbf{W}^{\mathrm{T}}\otimes\mathbf{U}\right)vec\{\mathbf{V}\} \quad (7)$$

$$vec\{\mathbf{U\Lambda W}\} = \left(\mathbf{W}^{\mathrm{T}}\odot\mathbf{U}\right)vecd\{\mathbf{\Lambda}\} \quad (8)$$

and, assuming $\mathbf{U}$ is $\alpha\times\beta$ and $\mathbf{W}$ is $\gamma\times\delta$:

$$vec\{\mathbf{UW}\} = (\mathbf{I}_\delta\otimes\mathbf{U})vec\{\mathbf{W}\} = \left(\mathbf{W}^T\otimes\mathbf{I}_\alpha\right)vec\{\mathbf{U}\} \quad (9)$$

$$vec\{\mathbf{U}\otimes\mathbf{W}\} = \left[\mathbf{I}_\beta\otimes\mathbf{K}_{\delta,\alpha}\otimes\mathbf{I}_\gamma\right]\left(vec\{\mathbf{U}\}\otimes vec\{\mathbf{W}\}\right) \quad (10)$$

ACKNOWLEDGMENT

Work of I. Kopriva has been partially supported through the FP7-REGPOT-2012-2013-1, Grant Agreement Number 316289 – InnoMol.

6. REFERENCES

[1] I. Kopriva, and A. Cichocki, "Blind decomposition of low-dimensional multi-spectral image by sparse component analysis," *Journal of Chemometrics*, vol. 23, pp. 590-597, 2009.

[2] T. C. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A Convex Analysis-Based Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing," *IEEE Tr. Sig. Proc.*, vol. 57, pp. 4418-4432, 2009.

[3] I. Kopriva, and A. Cichocki, "Blind multi-spectral image decomposition by 3D nonnegative tensor factorization," *Optics Letters*, vol. 34, pp. 2210-2212, 2009.

[4] Q. Zhang, H. Wang, R. J. Plemons, and V. P. Pauca, "Tensor methods for hyperspectral data analysis: a space object material identification study," *J. Opt. Soc. Am. A*, vol. 25, pp.3001-3012, 2008.

[5] T. Nakai, S. Muraki, E. Bagarinao, Y. Miki, Y. Takehara, K. Matsuo, C. Kato, H. Sakahara, and H. Isoda, "Application of ICA to magnetic resonance imaging for enhancing the contrast of gray and white matter," *Neuroimage*, vol. 21, pp. 251-260, 2004.

[6] M. J. McKewon, T. P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T. W. Lee, and T. J. Sejnowski, "Spatially independent activity patterns in functional MRI data during the Stroop color-naming task," *Proc. Natl. Acad. Sci. USA*, Vol. 95, pp.803-810, 1998.

[7] X. Hu, A. Shimizu, H. Kobatake, and S. Nawano, "Independent Component Analysis of Four-Phase Abdominal CT Images," *LNCS* 3217, pp. 916-924, 2004.

[8] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*. John Wiley, Chichester, 2004.

[9] J. D. Carrol, and J. J. Chang, "Analysis of individual differences in multidimensional scaling via N-way generalization of Eckart-Young decomposition," *Psychometrika*, vol. 35, 283-319, 1970.

[10] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an exploratory multi-mode factor analysis," *UCLA Working Papers in Phonetics* 16, pp.1-84, 1970.

[11] N. D. Sidiropoulos, and R. Bro, "On the uniqueness of multilinear decomposition of *N*-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229-239, 2000.

[12] P. Comon, "Tensors: a brief introduction", *IEEE Sig. Proc. Magazine*, vol.31, May, 2014, doi: 10.1109/MSP.2014.2298533.

[13] Y. D. Kim, A. Cichocki, and S. Choi, "Nonnegative Tucker Decomposition with Alpha-Divergence," in *Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, March 30-April 4, 2008, pp. 1829-1832.

[14] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations*. John Willey & Sons, 2009.

[15] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279-311, 1966.

[16] X. Q. Liu, and N. D. Sidiropoulos, "Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2074-2086, 2001.

[17] A. Hjørungnes. and D. Gesbert, "Complex-Valued Matrix Differentiation: Techniques and Key Results," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2740-2746, 2007.

[18] C. R. Rao, *Linear Statistical Inference and its Applications*, ser. Probability and Statistics. Wiley, 1965.

[19] H. Lev-Ari, "Efficient Solution of Linear Matrix Equations with Application to Multistatic Antenna Array Processing," *Comm. in Inf. and Syst.*, vol. 5, no. 1, pp.123-130, 2005.

[20] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. on Circuits and Systems*, vol. 25, no. 9, pp. 114-122, Sep. 1978.

[21] J. R. Magnus and H. Neudecker, "Matrix differential calculus with applications to simple, Hadamard and Kronecker products," *Journal of Math. Psychol.* vol. 29, no. 4, pp. 474-492, 1985.